Optimization of ATM and Legacy LAN for High Speed Satellite Communications[1]

Wendy R. Schmidt, Jeffrey L. Tri, Marvin P. Mitchell, Steven P.
Levens, Merrill A. Wondrow, Leslie M. Huie, Robert E. Martin,
Barry K. Gilbert, Bijoy K. Khandheria,
Mayo Foundation
Rochester, MN 55901

Abstract

A high data rate (HDR) terrestrial and satellite network was implemented to transfer medical images and data. This paper describes the optimization of the workstations and networking equipment. Topics include tuning the network software configuration of Sun Microsystems workstations, Fore Systems ATM switches, and Cisco routers, as well as the transfer rate results of four distinct telemedicine experiments.

The researchers were successful in achieving the transfer rates needed by the telemedicine software; particularly important was the proper determination of peak transfer rates and window sizes in making use of the resources available to the network interface cards (NICs) on the Sun Microsystems and Hewlett Packard workstations.

1. Introduction

Telemedicine is a unique way to offer a medical service to a patient who would otherwise not have access to medical specialists. Globally, interest in telemedicine continues to increase, particularly in the store-and-forward form wherein complex images and comprehensive patient data are recorded for later review by a specialist at a remote location. This form of telemedicine requires the use of a reliable, on-demand, high speed network to effectively handle the large quantities of data necessary to ensure an accurate diagnosis. Ultimately, development of a high speed network for telemedicine must include terrestrial and satellite links as well as ATM and legacy LAN technology. This type of high speed network could then accommodate both national and international demand for telemedicine. In the HDR experiments, Mayo Clinic was interested in demonstrating the clinical and technical feasibility of using a combined terrestrial and Advanced Communication Technology Satellite (ACTS) network to provide on-demand short duration access for telemedicine between remote hospitals and tertiary care centers. This project included four evolving areas of telemedicine: angiography, echo-cardiology, family medicine, and radiology. This paper focuses on the technical feasibility of implementing and tuning a network utilizing ATM, FDDI, and Ethernet.

The goal of the network support team was to design a network that provided a high speed communications link between medical workstations at multiple sites. The network was continuously monitored to verify proper operation and assist in performance evaluation. The network team sought to tune the network and workstations such that access to patient data appeared identical to physicians on both sides of the satellite link.

[1] The authors gratefully acknowledge the support of the Advanced Research Projects Agency, contract number DABT63-93-C-0027.

The contract required the network to incorporate both the ACTS satellite and the OC-12, 622 megabits per second (Mbps), MAGIC terrestrial backbone maintained by Sprint Corp. The technical team decided to purchase ATM networking equipment from a single vendor to minimize interoperability issues. Additionally, TCP/IP was selected as the data transfer protocol due to its reliability.

2. Network

This section presents the network architecture used in the experiments and introduces the network parameters that were central to the optimization discussed in Section 4.

2.1 Terminology

The following terms will be used throughout the paper:

- *Window size* is defined as the number of unacknowledged bytes of transmitted data allowed for a given connection between workstations.
- *Send space* is defined as the number of bytes in the TCP send buffer.
- *Receive space* is defined as the number of bytes in the TCP receive buffer.
- *Round trip time*, RTT, is defined as the time elapsed starting when a packet is sent and ending with the receipt of the acknowledgment for that packet. For example, “ping -s” between two workstations on either side of the satellite link reported a RTT of 542 ms.
- *Workstation tuning* refers to adjusting TCP parameters to maximizing transfer rates while minimizing network errors.
- *Data throughput* refers to the amount of data transferred between workstations in a given time frame.

2.2 Network Architecture

Figure 1 depicts the terrestrial and satellite OC-3 (155 Megabits per second) links among the five medical facilities involved in the experiments. The satellite link was a BPSK mode OC3 connection.

SPANS, Fore Systems’ proprietary version of switched virtual circuits (SVCs), was chosen to provide ATM connections between most workstations. SVCs provide self-healing network connections. However, SPANS only works on Fore Systems’ equipment. The alternative protocol for SVCs is defined by the ATM User-Network Interface Specification, UNI 3.0. UNI signaling provides interoperability between ATM equipment vendors but was not selected because it introduced the potential for a single point of failure at the server which controls address registration.

Cisco routers do not use SPANS. The traffic received by one Cisco router was always destined for the other Cisco router. Hence, traffic needed to be “tunneled” through the ATM network between the Cisco routers. Fore Systems’ smart permanent virtual circuits (SPVCs) provide this function. These SPVCs are set up on the ATM switches connected to the Cisco 7000 routers. A bi-directional PVC is configured between one router and the switch to which it is connected; identical PVCs are set up for the other router/switch pair. Communications between the switches is handled by SPANS. The inherent strength of SPVCs is that traffic is automatically re-routed through an alternate path by SPANS if one of the terrestrial links goes down. Essentially SPVCs provide the necessary self-healing robustness of SVCs between switches.

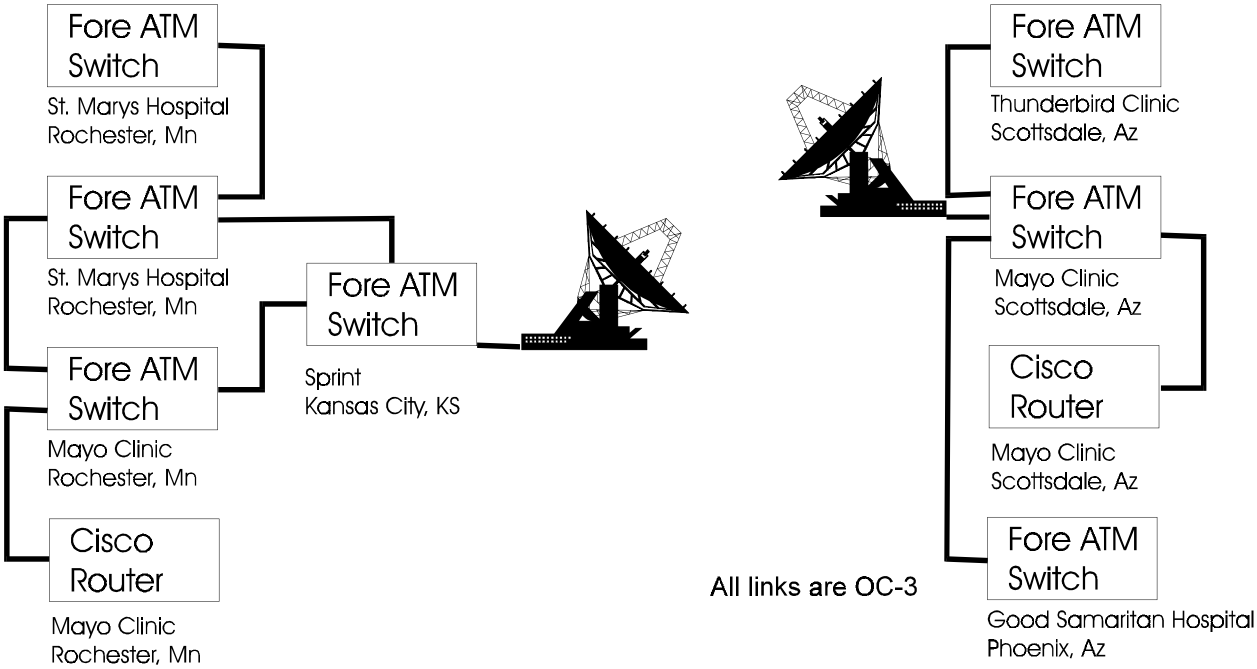


Figure 1: ATM and Legacy LAN Network Diagram

The legacy LAN was composed of four FDDI rings, three Ethernet segments, and two ATM interfaces (Figure 2). Sun and Hewlett Packard (HP) workstations running UNIX were present on the ATM network and legacy LAN as discussed in Section 3.

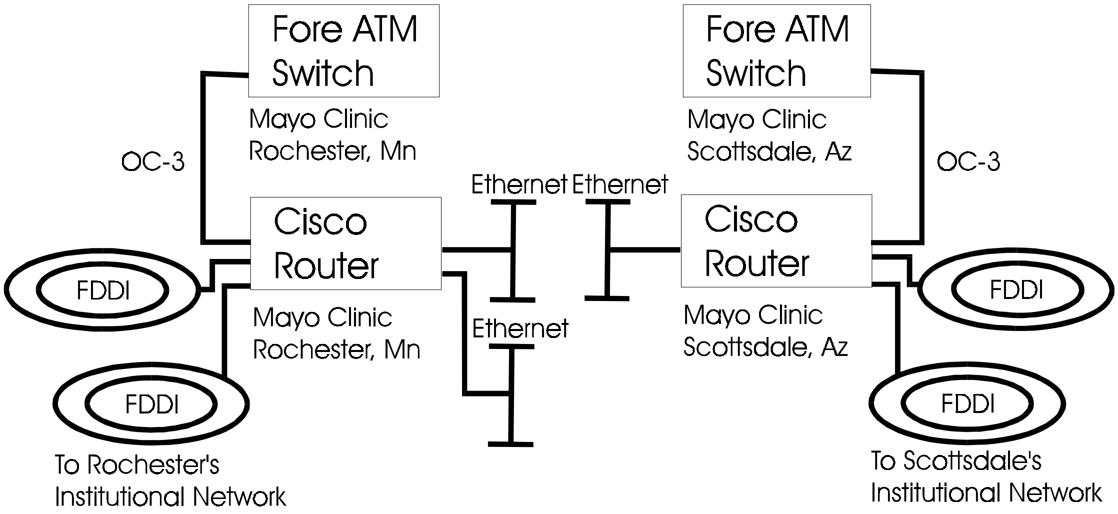


Figure 2: Legacy LAN

2.2.1 ATM Laboratory Configuration

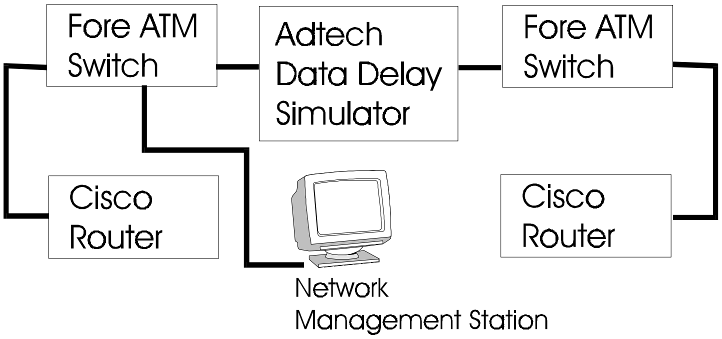


Figure 3: Lab Setup

The ATM laboratory at Mayo Clinic in Rochester, Minnesota was set up to simulate the network and satellite links to pre-test and tune the equipment and applications. There were two Fore Systems ASX-200 ATM switches in the laboratory, each representing one side of the satellite link (Figure 3). Adtech's data delay simulator, SX-14, simulated the satellite link by adding a round trip delay of 600 ms. A network management station running HPOpenview and ForeView was also installed in the lab to monitor the ATM network. ForeView is Fore Systems' ATM network management software.

2.3 Network Parameters

2.3.1 This section introduces four network configuration issues that were found to be key to the network optimization discussed in Section 4: TCP Extensions for High Performance, TCP parameters and MTU size, and source quenching. Other TCP parameters were not found to have a significant effect on the transfer rates in these experiments.

2.3.2 CP Extensions for High Performance [1]

The maximum window size in standard TCP/IP is 64 Kbytes. This translates to a maximum transfer rate of less than 1 Mbps between workstations communicating over the satellite link when the RTT is 542 ms (see Equation 1). To transfer data at speeds greater than 1 Mbps, a larger maximum window size is required.

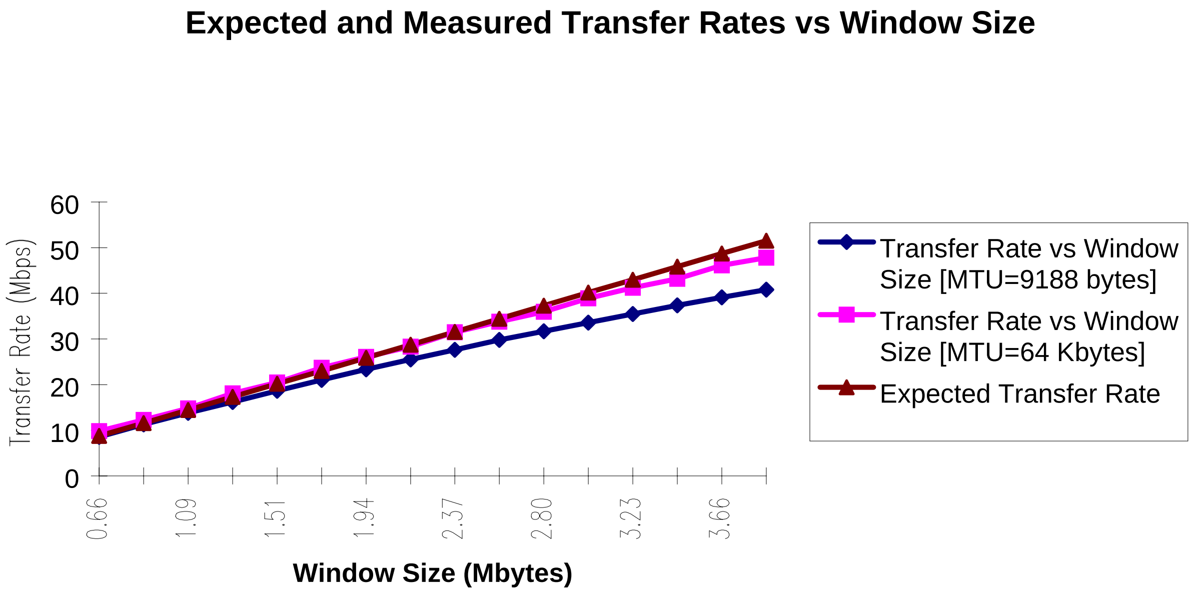
Equation 1: Throughput (Mbits/sec) = [Window size (Mbytes) * (8 bits/byte)] / RTT

RFC 1323, TCP Extensions for High Performance, proposes extending the maximum window size from 64 Kbytes (a 16 bit value) to 2^{30} bytes (a 32 bit value). Software packages that implement RFC 1323 were installed on all of the Sun workstations communicating over the satellite link. This software was used extensively by the technical team in the lab and during the experiments to optimize transfer rates over the satellite link. In theory, workstations could communicate at OC-3 rates over the satellite link simply by changing the send and receive window sizes. To obtain the desired throughput of 155 Mbps with an RTT of 542 ms, one would simply by increasing the send and receive window sizes to 10.5 Mbytes (see Equation 2).

Equation 2: Window size (Mbytes) = [Desired throughput (Mbits/sec) * RTT] / (8 bits/byte)

2.3.3 TCP Parameter Settings and MTU Size

Adjusting the window size and MTU size are the basic means for controlling data transfer rate. The reasons for restricting the maximum transfer are discussed in Section 4.1. The near linear relationship between throughput and window size with a RTT of 601 ms is shown in Graph 1. Changing the MTU size from the default 9188 bytes to the maximum of 64 Kbytes[2] on ATM hosts increased the data throughput as shown in Graph 1.



Graph 1: Memory to Memory Data Throughput vs Window Size (RTT = 601 ms)

The maximum value of the TCP congestion window, which defaults to 0.25 Mbytes, clips the achievable data throughput when it is smaller than the window size. Hence, the maximum congestion window should be equal to or larger than the desired window size.

2.3.4 Source Quenching

In source quenching, the sender is configured to transmit at less than the link's maximum of 155 Mbps. This decreases the potential for overrunning the available resources on the receiver. In the telemedicine experiments where source quenching is necessary, it is performed by establishing permanent virtual circuits with peak rates appropriate to each application as discussed in Section 4.

3. Experiment Overviews

Section 3 provides an overview of the four telemedicine experiments forming this project between the Arizona and Minnesota sites: angiography, echo-cardiology, family medicine, and teleradiology. A brief description is given from the physician's point of view along with the technical elements that make each experiment distinct.

[2] A workaround that permitted a larger maximum MTU size for the angiography workstations is discussed in Section 4.1.

3.1 Angiography

The angiography experiment was set up to accommodate both intra-procedure and post-procedure consultations. An intra-procedure consultation is conducted during the angiographic study. While the patient undergoes coronary angiography at one of the clinical sites, the images are collected digitally and immediately transmitted across the HDR satellite network. The images are therefore viewed during the procedure by physicians at both sites. In a post-procedure consultation, the images are stored (either digitally or on cine film) and the patient is then discharged from the cardiac laboratory. To perform the consultation, cine films (if used for storage) are digitized. The digital images are transmitted to the remote site, again via the HDR satellite network, and reviewed simultaneously by physicians at both sites at some later time.

The angiography experiment involved six ATM hosts: a local and remote cardiac review station, a local and remote archive manager, and a local and remote gateway (Figure 4). The details of these workstations and the TCP configuration of the archive managers are available in Appendix B. The digital tape library and 12 gigabyte RAIDs (redundant array of independent disks) provided storage of the patient images transferred from the catheterization laboratory. Images were also stored on cine film; these films were digitized and transferred to the archive manager.

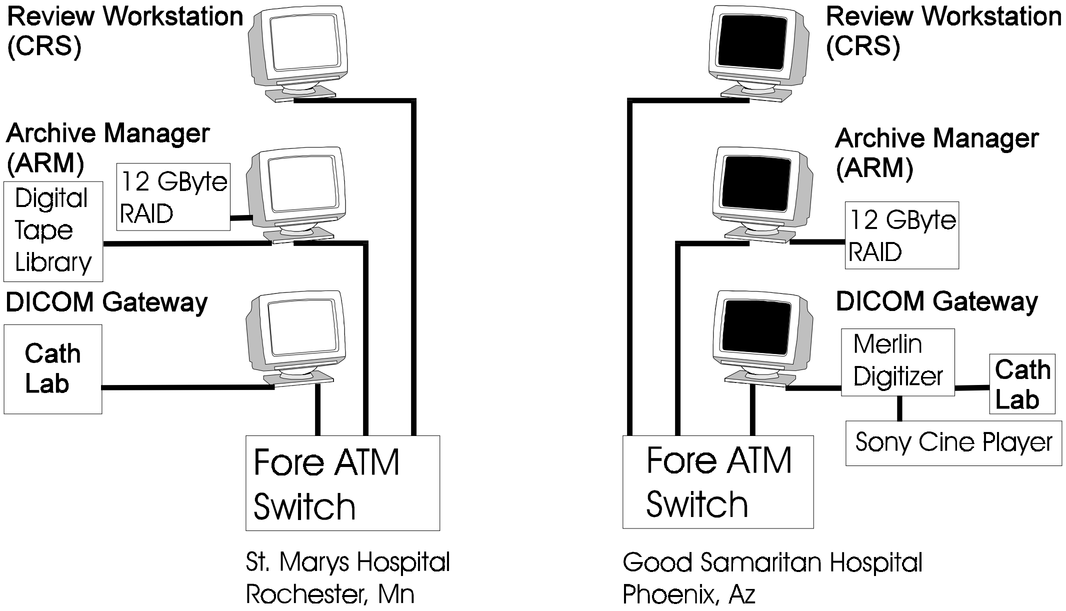


Figure 4: Angiography ATM Network

3.2 Echo-cardiology

As in angiography, the echo-cardiology experiment involved intra-procedure and post-procedure components. In the intra-procedure portion of the experiment, live video or echocardiograms from an HP ultrasound machine were transferred across the satellite link. The post-procedure portion utilized a VCR or a Magneto Optical Disk (MOD) as the source of echocardiograms to transmit. Only one source could be transferred at a time.

This experiment involved two ATM hosts, one on either end of the satellite link. The host in Arizona had connections to the HP ultrasound machine and VCR. There were MOD storage device drives on both workstations. The Minnesota site was the origin for the live video source.

3.3 Family Medicine

The family medicine experiment utilized the store-and-forward concept. Patient data was digitized and stored on a server in Arizona for later transmission over the satellite link. Specialists at the remote site in Minnesota accessed the consultations at their leisure.

This experiment utilized six Sun workstations connected to the ATM network. A server which controlled the transfer of patient data was located on either side of the satellite link. The remaining four workstations were used to download the patient consultations from their local server.

3.4 Teleradiology

Magnetic resonance (MR) and computed tomography (CT) images were collected in Minnesota and transferred across the satellite link for later review by a radiologist at the remote site in Arizona. The teleradiology experiment included thirteen Sun workstations on either FDDI or Ethernet. The eleven workstations in Rochester were responsible for collecting and transmitting MR and CT images. In Arizona, the images were received and reviewed on the remaining workstations.

4. Network Optimization and Limiting Factors

Section 4 discusses the tuning parameters found to be key in achieving maximum network performance and highlights the factors limiting transfer rates beyond the fundamental limitations of the workstation processors. This section is organized by the experiments in which the tuning parameters and limiting factors were most evident.

4.1 Angiography

The angiography experiment highlighted two issues: source quenching and connection establishment. Note that to efficiently handle the 0.25 Mbytes of data per frame of cine film, the maximum MTU size was established during boot up of the Sun workstations to be 0.25 Mbytes.

4.1.1 Source Quenching

There were three instances in which the network support team observed cells being dropped by receiving workstations in Arizona. These dropped cell errors occur due to insufficient resources available to the network interface card (NIC). This occurred when data was transferred from the gateway to the archive manager, between archive managers, and from the archive manager to the cardiac review station.

Simply reducing the window size can stop the errors but does not yield the maximum achievable transfer rate. Maximizing the actual transfer rate without cells being dropped at the receiver requires establishing a peak rate for the connection as well adjusting the TCP window size. A peak rate can be established when a permanent virtual circuit (PVC) is established. An outer iteration on the peak rate set for the PVC and an inner iteration on the TCP window size were performed as follows:

1. Select a peak rate for the PVC
a. Iterate on the TCP window size to determine the maximum window size which does not yield errors (as reported by the NIC) [3].
b. Observe the actual data transfer rate.
2. Repeat step 1 with increasing peak rates until the actual transfer rates no longer increase.

When the iterations are complete, the actual transfer rate is being limited by the resources available to the NIC instead of the peak rate on the PVC.

4.1.2 Connection Establishment

In the angiography experiment, the gateway receives data at 60 Mbps (30 frames per second) from the Merlin digitizer. It must immediately transfer this data to the archive manager to avoid running out of resources. PVCs were set up to avoid the delay present with SVCs in establishing a connection.

4.2 Echo-cardiology

The echo-cardiology experiment revealed limitations in the physical storage medium and in an implementation of TCP extensions for high performance.

4.2.1 TCP Extensions for High Performance

The Hewlett Packard workstations had a software patch available that implemented RFC 1323. With this patch, the maximum size of the send and receive windows could be increased from 64 Kbytes to 868 Kbytes (limiting the maximum throughput to 13 Mbps) without running out of resources.

4.2.2 Physical storage media

A MOD storage device drive was connected to an echo-cardiology workstation on either side of the satellite link. The physics of reading and writing data to the MOD drives limited transfer rates to less than 3 Mbps, as compared to the approximately 20 Mbps transfer rate achievable when using an internal SCSI drive.

4.3 Family Medicine

The family medicine experiment encountered the resource limitations on the NICs as discussed in Section 4.1.1. The transfer rates were improved by increasing the MTU size as illustrated in Section 2.3.2.

Difficulties with application software were observed in the database program which handled the synchronization of the patient data in the servers on either side of the satellite link. The database program utilized only the default TCP parameters when establishing a connection and transferring data. To use the non-default parameters which were required for higher throughput, FTP was used to perform transfer tasks for the database software.

[3] If there are still errors reported by the receiver, it may be necessary to reduce the maximum congestion window parameter as discussed in Section 2.3.2.

4.4 Teleradiology

The teleradiology experiment revealed three issues: interoperability between ATM equipment providers, window size interaction with workstation's FDDI NIC, and the file size.

4.4.1 Interoperability

In a standard configuration, the Cisco router has its SONET level timing slave to the timing of the ATM switch. However, this configuration on the Cisco 7000 router with the Fore Systems' ASX-200 ATM switch yielded the loss of thirty to fifty percent of packets being transmitted. This problem was resolved by configuring both pieces of equipment to use their own internal timing.

4.4.2 Window Size

There were problems associated with configuring large window sizes (greater than approximately 1 Mbyte) on the legacy LAN hosts. The FDDI hosts reported an unacceptable number of output errors as the window size was increased. In normal operation, the acceptable number of output errors is low: 0.025% of the total number of output packets [2]. The window size was lowered until no output errors were reported by the Sun workstation.

4.4.3 File Size

In both the FDDI and ATM hosts, it was noted that the average transfer rate increased with file size. It is believed that the slower average transfer rates reported when transferring small files was due to the overhead associated with the TCP slow start mechanism and with the setting up and tearing down of a socket [3]. For each file to be transferred by FTP a socket is set up, the data is transferred, then the socket is torn down. This sequence is repeated until all the requested files are transferred. It was estimated that there were 30 seconds of overhead with each socket setup on a network with a RTT of 600 ms.

4.5 Satellite Availability

The intermittent availability of the satellite link was another issue for consideration. It introduced another level of complexity into the design and management of the networking equipment and applications. For example, SPVCs would have allowed the establishment of peak transfer rates on the ATM hosts while maintaining the self-healing robustness of SVCs. However, this was not a viable option because the intermittent availability of the satellite link caused the ATM switches through which the hosts were communicating to "lock up" requiring a manual reset of the switch.

4.6 Summary

In telemedicine, the selection of hardware must accommodate both the medical applications and telemedicine's demand for high data throughput. Given the limitations of this hardware there remain a number of network configuration issues to be resolved to achieve the desired high data throughput. Particularly interesting among these issues was the determination of the peak connection data rate and the corresponding window sizes that maximized the usage of resources available to the NIC.

5. Experimental Results

Table 1 contains the peak data transfer rates over the satellite link measured after the network and workstation optimization discussed in Section 4.

Table 1: Post Optimization Transfer Rates Per Experiment

Experiment	Type of Data	Average File Size	Window Size	Calculated Transfer Rate (Equation 1)	Measured Transfer Rate
Angiography (Post-procedure)	Digitized Cine Film 1 Study (3 - 300 Frame Sequences)	225 MB	1.5 MB	23.2 Mbps	21 Mbps
Angiography (Intra-procedure)	1 sequence - 300 frames	75 MB	1.5 MB	23.2 Mbps	21 Mbps
Echo-cardiology (analog video)	digitized video	N/A	868 KB	13.12 Mbps	11.7 Mbps
Echo-cardiology (magneto optical disk)	Digitized Echo-cardiology Studies (200 loops / side)	1.25 MB / 1 loop	868 KB	13.12 Mbps	2.8 Mbps
Telemedicine: Cardiology	Angiography, Echo-cardiology, and Digitized Chest Xrays	300 MB	1.5 MB	23.2 Mbps	20.2 Mbps
Telemedicine: Dermatology	Video File	250 MB	1.5 MB	23.2 Mbps	20.2 Mbps
Telemedicine: Orthopedic	Digitized Xrays	10 MB	1.5 MB	23.2 Mbps	20.2 Mbps

Table 2 contains an example of the effects of cumulative optimization. It reports transfer rates achieved in the angiography experiment for three different connection types at each of three steps in the optimization process. Two of the connection types are local and one is across the satellite link. All three cases show a significant improvement in transfer rate following the optimization process.

Table 2: Transfer Rates Original and Post-optimization

Workstations	Original Peak Transfer Rate	Peak Transfer Rate with Workstation Optimization	Peak Transfer Rate with Workstation and Network Optimization
Archive Manager to Archive Manager	1.1 Mbps (0.5 fps)	12 Mbps (6 fps)	21 Mbps (10 fps)
Archive Manager to Cardiac Review Station	8 Mbps (4 fps)	12 Mbps (6 fps)	19 Mbps (9.5 fps)
Gateway to Archive Manager	10 Mbps (5 fps)	21 Mbps (10.5 fps)	47.5 Mbps (24 fps)

6. Conclusion

The four telemedicine experiments discussed in this paper were implemented using ATM, FDDI, and Ethernet technology. The central factor in these experiments was a need for a high speed ATM network across the satellite link. This paper has described the difficulties that were encountered in implementing / configuring this network for use with the telemedicine software. In particular, the proper tuning of peak transmission rates and window sizes allowed the successful transmission of the patient data including the 21 Mbps transmission rate of the angiography studies in the angiography experiment.

Acknowledgments

The authors thank: Mr. Bill Eden’s of General Electric for his design and support of the angiography workstations; Mr. Bern Stump from HP for his design and support of the echo-cardiology workstations; and Mr.Vince Kluth from InfoHealth for his support of the family medicine experiment.

References

1. Jacobson, V., R. Braden, and D. Borman. *TCP Extensions for High Performance*. LBL 1992; Internet RFC 1323.

2. Sun Microsystems Frequently Asked Questions, “Data Corruption on the Network,” Document ID: 0928, Aug. 23, 1995.

3. W. R. Stevens. *TCP/IP Illustrated, Volume 1: The Protocols*. Addison-Wesley, Reading, Massachusetts, 1994.

Appendix A
TCP parameter descriptions

The following TCP parameter definitions are drawn from Sun workstation documentation for tcp-lfn.

tcp_tstamp_always: If this parameter is nonzero, a timestamp option will always be sent when connecting to a remote system. The default is zero.

tcp_max_buf: This parameter specifies the maximum buffer size a user is allowed to specify with the SO_SNDBUF or SO_RCVBUF options. Attempts to use larger buffers will fail with EINVAL. The default is 256K (262144). Note that it is unwise to make this parameter much larger than the maximum buffer size your applications require, since that could allow malfunctioning of malicious applications to consume unreasonable amounts of kernel memory.

tcp_xmit_hiwat: This parameter specifies the default value for a connection's send buffer space; that is, the amount of buffer space allocated for sent but unacknowledged data. The default is 8K. In most cases, tcp_xmit_hiwat should be identical to tcp_recv_hiwat.

tcp_recv_hiwat: This parameter specifies the default value for a connection's receive buffer space; that is, the amount of buffer space allocated for received data (and thus the maximum possible advertised receive window). The default is 8K. In most cases, tcp_recv_hiwat should be identical to tcp_xmit_hiwat.

tcp_host_param: This parameter is actually a table of IP addresses, networks, and subnetworks, along with default values for certain TCP parameters to be used on connections with the specified hosts. The table may be displayed with ndd as follows:

```
example# ndd /dev/tcp tcp_host_param
Hash      HSP                 Address                       Subnet Mask   Send              Receive
          Tstamp
027       fc31eea4            129.176.238.131               000.000.000.000         0000128000    0000128000
          1
```

Each element in the table specifies either a host, network (with optional subnet mask), or subnet, along with the default send buffer space and receive buffer space, and a flag indicating whether timestamps are to be used.

The default values specified in the table are used for both active and passive connections (that is, both connect () and listen ()). The most applicable match found is used; first the full host address, then the subnet, and finally the network.

The example table above specifies that:

Connections with the host 129.176.238.131 will use send and receive buffer sizes of 128000 bytes, and will use timestamps.

The send and receive space values from the tcp_host_param table will only be used if they are larger than the values set by the user (or obtained from tcp_xmit_hiwat and tcp_recv_hiwat). This is so that the user can specify larger values for improved throughput and not have them erroneously reduced.

For the Mayo experiments, only specific host addresses were used and Tstamp was always set to 1.
The results from the Mayo experiments indicate that the send and receive space values from tcp_host_param are ALWAYS used even when tcp_xmit_hiwat and tcp_recv_hiwat are larger than the send and receive space values.

tcp_cwnd_max: Slow start adds the congestion window to the sender's TCP. When a new connection is established with a host on another network, the congestion window is initialized to one segment (i.e., the segment size announced by the other end). Each time an ACK is received, the congestion window is increased by one segment. (*cwnd* is maintained in bytes, but slow start always increments it by the segment size.) The sender can transmit up to the minimum of the congestion window and the advertised window. The congestion window is flow control imposed by the sender, while the advertised window is flow control imposed by the receiver. [3, p. 285)

Appendix B

Angiography Workstation Details

The angiography workstations are typical of the workstations used in the four telemedicine experiments:

Host	Workstation Type	Processor Speed	RAM	Operating System	ATM NIC
Az ARM	Sun Sparc 20	Dual 50 Mhz	132 MB	Solaris 2.4	SBA-200
Mn ARM	Sun Sparc 10	50 Mhz	132 MB	Solaris 2.4	SBA-200E
Az CRS	HP 725	100 Mhz	132 MB	HP-UX 9.05	ESA-200
Mn CRS	HP 735	100 Mhz	148 MB	HP-UX 9.03	ESA-200
Az Gateway	Sun Sparc 5	110 Mhz	132 MB	Solaris 2.3	SBA-200
Mn Gateway	Sun Sparc 5	85 Mhz	132 MB	Solaris 2.3	SBA-200E

Table B.1

The angiography workstations used the following TCP parameters to tranfer data over the satellite link between the Az and Mn ARMs detailed in the table above.

TCP parameter	Default Value	Satellite Transfer: Sun to Sun
Time stamp	0	1
Maximum buffer size	256 Kbytes	15 Mbytes[4]
SO_SNDBUF	8192	1.5 Mbytes
SO_RCVBUF	8192	1.5 Mbytes
Congestion Window	256 Kbytes	1.5 Mbytes

Table B.2

[4] Maximum buffer size in the other experiments was equal to max (SO_SNDBUF, SO_RCVBUF) which is significantly less than the 15 Mbytes required by angiography application.